

Initiation à la méthodologie de la recherche clinique

Module 4 : Les études cas - témoins

BEKER 2025/2026

Professeur Samira ABROUK

Etude cas - témoins

Protocole

- Les études cas-témoins (case-control studies) sont des études épidémiologiques observationnelles, analytiques, étiologiques, dont le principe consiste à comparer la fréquence d'une exposition passée à un ou plusieurs facteurs d'exposition (facteurs de risque potentiels) entre 2 groupes de sujets:
- **les cas**: sujets atteints de la maladie étudiée
- **les témoins**: sujets indemnes de la maladie étudiée
- **L'exposition à un facteur précède la survenue de la maladie**
- Par définition, **l'étude cas-témoin** est une étude **rétrospective**.

Etude cas - témoins

Identification des FR ou des FP

- L'identification des facteurs de risque ou des facteurs protecteurs se fait à l'aide de l'Odds Ratio.

	Malade	Non malade
Exposé	a	b
Non exposé	c	d

$$OR = \frac{ad}{bc}$$

- On commence par calculer les OR bruts liés à chaque facteur d'exposition avec leur intervalle de confiance à 95%.
- Pour évaluer la part propre de chaque facteur, on calcule l'OR ajusté à partir du modèle logistique.

Etude cas - témoins

Calcul de la taille minimale de l'échantillon

1- Autant de cas que de témoins

- Nombre de cas : n

$$n = \frac{(z_{\alpha} \sqrt{2pq} + z_{\beta} \sqrt{p_1 q_1 + p_0 q_0})^2}{(p_1 - p_0)^2}$$

Avec

p_0 = Fréquence d'exposition dans la population non malade

p_1 = Fréquence d'exposition attendue dans la population malade

$$p_1 = \frac{p_0 OR}{1 + p_0 (OR - 1)} \quad \text{et} \quad p = \frac{p_1 + p_0}{2}$$

Etude cas - témoins

Calcul de la taille minimale de l'échantillon

2- un cas pour k témoins

$$n = \frac{\left[z_{\frac{\alpha}{2}} \sqrt{\left(1 + \frac{1}{k}\right) p' q'} + z_{\beta} \sqrt{p_1 q_1 + p_0 q_0} \right]^2}{(p_1 - p_0)^2}$$
$$p' = \frac{(p_1 + k p_0)}{1 + k}$$

Avec

n correspond au nombre de cas et k*n au nombre de témoins

- **Attention! Il est inutile d'avoir $k > 4$**
- **On ne gagne rien sur le nombre de cas et on augmente le nombre total de sujets**

Etude cas - témoins

Analyse d'une étude cas-témoins

1- Analyse facteur par facteur

- Calcul de l'OR
- L'OR doit être exprimé avec son intervalle de confiance à 95%
- Le risque n'est significatif que si l'IC ne contient pas la valeur 1
- 3 possibilités:
 - L'IC ne contient pas la valeur 1 et tout l'intervalle se situe avant 1: par exemple (0,20 0,40): **Facteur protecteur significatif**
 - L'IC ne contient pas la valeur 1 et tout l'intervalle se situe après 1: par exemple (1,58 3,43): **Facteur de risque significatif**
 - L'IC contient la valeur 1: **Facteur non significatif**

Etude cas - témoins

Analyse d'une étude cas-témoins

2- Analyse multivariée

Toutes les variables explicatives significatives ou à la limite de la significativité ($p \leq 0,10$), retrouvées précédemment, sont introduites dans un modèle logistique pour évaluer la part propre de chaque facteur.

On peut utiliser plusieurs types de modèles:

- Le modèle ascendant: Le modèle retenu est alors le modèle à k variables. Les variables sont introduites pas à pas, la 1^{ère} introduite est celle qui est la plus significative
- La 2^{ème} variable (qui arrive juste après d'un point de vue significativité) n'est retenue dans le modèle que si la vraisemblance du modèle avec 2 variables est plus élevée significativement que le celle du modèle à 1 variable.
- On continue à introduire les variables pas à pas jusqu'à ce qu'il n'y ait plus de différence significative entre le modèle à k variables et le modèle à k+1 variables. Le modèle retenu est alors le modèle à k variables.
- Le modèle descendant: les n variables sont toutes introduites au début, on calcule la vraisemblance du modèle
- Puis la variable la moins significative est retirée; ce modèle n'est retenu que si la vraisemblance du modèle avec 2 variables n'est pas inférieure significativement que le celle du modèle avec toutes les variables,
- On continue à retirer les variables pas à pas jusqu'à ce qu'il y ait une différence significative entre le modèle à k variables et le modèle à k-1 variables. Le modèle retenu est alors le modèle à k variables.
- Le modèle « Entrée »: les variables sont toutes introduites et bloquées quelle que soit leur significativité

Etudes cas-témoins

Interprétation des données

Dans le modèle retenu

- Seront estimés les OR des facteurs considérés significatifs, avec leur intervalle de confiance à 95%
- Il s'agit d'Odds Ratio ajustés sur les autres facteurs, exprimant ainsi le risque propre du facteur dans la survenue de la maladie

■ Probabilité p pour un patient d'avoir la maladie

- Calculé à partir du modèle logistique:
- $$\text{Logit } p = \ln \frac{p}{1-p} = \alpha + \beta_1 X_1 + \dots + \beta_k X_k = a$$
- On en déduit :
$$p = \frac{e^a}{1+e^a}$$

Etudes cas-témoins

Critères de qualité d'un modèle logistique

- Pour valider la qualité du modèle logistique, on utilise :
 - **Le test de Hosmer-Lemeshow** qui compare les valeurs attendues à partir du modèle logistique avec les valeurs observées. Il ne faut pas qu'il y ait une différence significative entre le modèle théorique et les données observées.
 - **Le % de bien classés** parmi les malades et les non malades
 - **La courbe ROC** utilisée particulièrement dans la prédiction. Pour chaque sujet, la probabilité de survenue de la maladie étudiée est calculée en tenant compte des niveaux des facteurs de risque.
 - Les probabilités sont classées par ordre croissant et jouent le rôle d'une des valeurs quantitatives d'un test diagnostique. Pour chaque probabilité (seuil du test), la sensibilité et la spécificité sont calculées.
 - La courbe ROC représente la Se en fonction de 1-Sp.
 - Un bon modèle a une courbe ROC $\geq 0,80$
 - La probabilité (seuil), retenue dans la prédiction de la maladie, est celle qui correspond au seuil optimal de la courbe ROC.